

Key Word Use in Letters of Recommendation for Ophthalmology Residency Applicants According to Race, Gender, and Achievements

Sahil Aggarwal, BS¹ Seanna Grob, MD, MAS² Dhruva Banerjee, BA³ Preston J. Putzel, BS⁴
Jeremiah Tao, MD, FACS²

¹School of Medicine, University of California, Irvine, California

²Division of Oculofacial Plastic and Orbital Surgery, Department of Ophthalmology, Gavin Herbert Eye Institute, University of California, Irvine, California

³Department of Neurobiology and Behavior, University of California, Irvine, California

⁴Donald Bren School of Information and Computer Sciences, University of California, Irvine, California

Address for correspondence Jeremiah Tao, MD, FACS, Division of Oculofacial Plastic and Orbital Surgery, Department of Ophthalmology, Gavin Herbert Eye Institute, Irvine School of Medicine, University of California, 850 Health Sciences Road, Irvine, CA 92697 (e-mail: jeremiat@uci.edu).

J Acad Ophthalmol 2018;10:e163–e171.

Abstract

Objectives To identify differences in letters of recommendation (LORs) of applicants to a single ophthalmology residency program by gender, race, academic performance, and match outcome.

Design This was a retrospective analysis of LORS for 2,523 applicants (7,569 letters) to the University of California, Irvine ophthalmology residency program from 2011 to 2018.

Methods Programming scripts were employed to determine the number of times 22 key words from four thematic categories (standout words, ability, grindstone, and compassion) appeared in LORs for each applicant. A chi-square test was performed to assess for possible differences in the presence of each key word by the following characteristics: gender, underrepresented minority (URM) status, Alpha Omega Alpha (AOA) membership, the United States Medical Licensing Exam (USMLE) Step 1 score, and match outcome. Linear regressions were created to determine the frequency at which words in each thematic category appeared according to the same baseline characteristics.

Results In the LORs, females were more likely to be described as “empathetic” ($p = 0.002$), URMs were more likely to be described as “caring” ($p = 0.002$), high Step 1 scorers (≥ 240) were more likely to be described as “outstanding” ($p = 0.002$), and matched students were more likely to be described as “exceptional” ($p = 0.001$), “outstanding” ($p < 0.001$), and “superb” ($p = 0.001$). Standout words appeared more often in the LORs of AOA members, matched candidates, and high Step 1 scorers ($p < 0.001$ for all comparisons). “Competent” appeared more commonly in LORs for low Step 1 scorers ($p < 0.001$) and unmatched applicants ($p = 0.001$).

Conclusion This study identifies differences in LORs by gender, URM status, and achievement including successful ophthalmology residency match. Females and URMs were more likely to be described as “empathetic” and “caring,” respectively; otherwise,

Keywords

- ▶ letters of recommendation
- ▶ ophthalmology
- ▶ bias
- ▶ gender
- ▶ underrepresented minority
- ▶ Alpha Omega Alpha

received
July 20, 2018
accepted after revision
October 11, 2018

DOI <https://doi.org/10.1055/s-0038-1675842>.
ISSN 2475-4757.

Copyright © 2018 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
Tel: +1(212) 584-4662.

License terms



we detected no gender or racial disparities in key word use in LORs. Candidates with high USMLE Step 1 scores or AOA membership had a higher frequency of standout words in their LORs. Whether they were truly more qualified in various dimensions or if they benefited from a halo effect bias warrants further investigation. There was a significant difference in the number of standout words in LORs between matched and unmatched applicants, suggesting that key word frequency may be a relevant metric for LOR appraisal.

Securing an ophthalmology residency position in the United States is a highly selective process.^{1,2} Factors commonly cited as contributors to matching include the United States Medical Licensing Exam (USMLE) scores, grades, research experiences, academic achievements, Medical Student Performance Evaluations (MSPEs, also known as the dean's letter), and letters of recommendation (LORs).^{3,4}

LORs provide important supplementary information about applicants, such as personality traits, professionalism, and interpersonal skills. However, the usefulness of LORs in the resident selection process has been questioned due to a lack of standardization of content. As part of a 3-year pilot program introduced by the Association of University Professors of Ophthalmology (AUPO), letter writers in the 2018 application cycle had the option to complete a standardized LOR form for candidates as an alternative to a formal LOR.⁵ This LOR requires writers to rank an applicant in comparison to peers for several competencies.⁶

In addition to lack of standardization, LORs may suffer from bias that may disadvantage some groups. Studies conducted in different academic settings identified significant differences in LOR content by gender.⁷⁻¹⁰ In a large retrospective study of 6,000 applicants to 16 Yale residency programs, Ross et al identified differences in key words such as "exceptional," "best," and "outstanding" in MSPEs by race and gender.¹¹ Isaac et al identified similar biases in MSPEs of diagnostic radiology applicants.¹²

Besides gender and race, other biases in residency applications may create an unlevel playing field in the match process. USMLE Step 1 performance and Alpha Omega Alpha (AOA) academic honor society membership have been shown to carry disproportionate weight in selection.⁴ The extent to which these academic factors influence LOR content has not been studied.

The halo effect is a cognitive bias in which an assessment of a person, place, or thing assumes ambiguous information from concrete information.^{13,14} High academic achievers may also have strong interpersonal and professionalism qualities; however, the senior author (J. T.), who has read LORs for over a decade, hypothesizes that through the halo effect, candidates with high board scores receive positive descriptive words for competencies beyond what written examinations assess. AOA membership may similarly confer additional praise through this heuristic that may be especially common for medical students who often have very limited time and context with their letter writers.

In this study, we evaluate the ophthalmology residency LORs of candidates to a single institution over eight application cycles to explore differences in key word appearance and frequency by gender, underrepresented minority (URM) status, academic achievements, and match outcome.

Materials and Methods

This retrospective cohort study was granted exempt status by the University of California, Irvine (UCI) Institutional Review Board; therefore, no informed consent was required. From 2011 to 2018, a total of 2,524 allopathic (MD) and osteopathic (DO) candidates applied to the UCI ophthalmology residency program and were reviewed for this study. Applicants who did not provide USMLE Step 1 scores were removed from analyses. Applications of those who were reapplying after a prior year(s) unsuccessful match were included in analyses as independent applications due to presumed differences in LOR content by year of submission.

A schematic of our methods is provided in **Fig. 1**. Each application was exported from the SF Match as a Portable Document Format (PDF) file. Python (Python Software Foundation) was used to deidentify all applications (including identifiers for both the applicants and letter writers) and to write scripts for automated extraction of baseline characteristics (USMLE Step 1 and Step 2 scores, URM status, and AOA status). The 2018 match was the first time applicants reported URM status, and all statistical analyses associated with URM were confined to this group only.

At the time of application submission, AOA status was not reported for 1,092 applicants, meaning that they either did not have an AOA chapter at their school or their AOA status was not yet determined. To ensure an accurate depiction of the role of AOA status on letter content, only those applications with definitive reporting of yes AOA status ($n = 1,431$) at the time of application were included in the analyses. Adobe Acrobat Professional's Optical Character Recognition software (Adobe Systems Incorporated) was used to convert all scanned pages of each PDF (including LORs) into searchable text. For processing of the searchable LORs, several Python programming scripts were written to extract key words. Using methodology as previously described, 22 key words from four thematic categories were investigated (**Table 1**).^{11,15} These scripts reviewed each LOR line by line for the key words, and the output data were the number of times each word appeared in all three LORs for each

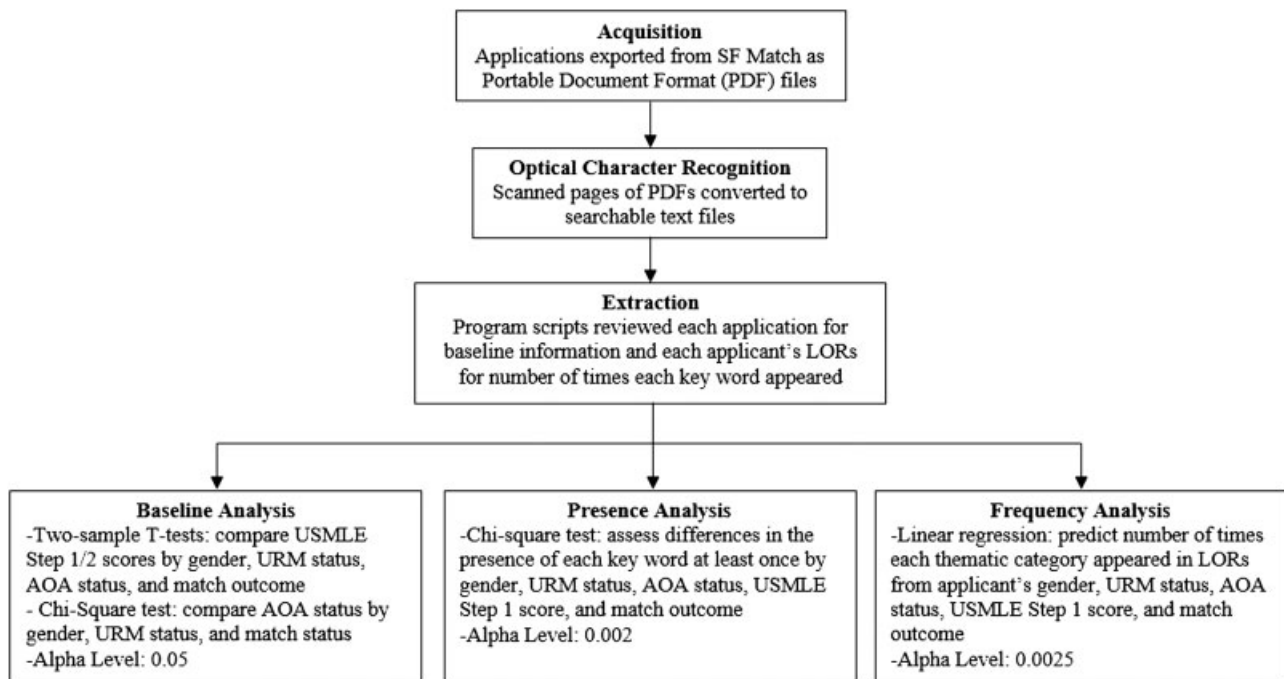


Fig. 1 Methods for key word analysis of letters of recommendation (LORs). AOA, Alpha Omega Alpha; URM, underrepresented minority; USMLE, United States Medical Licensing Exam.

applicant. The scripts also accounted for different forms of each key word, including “compassionate” (“compassion”), “diligent” (“diligence”), “empathy” (“empathetic”), and “talented” (“talent”). Because gender is not explicitly reported in each application, gender of each applicant was determined based on the frequency of the personal pronouns “he” and “she” in the LORs.

The final database was deidentified in Python and imported into the statistics package R (The R Foundation for Statistical Computing, Vienna, Australia) for analysis. Two-sample *t*-tests were completed to compare USMLE Step 1 and Step 2 scores by gender, URM status, AOA status, and match outcome (match or no match). A Pearson’s chi-square test was performed to compare AOA status by gender, URM status, and match status. An α level of 0.05 was used to assess for significance in these analyses.

Table 1 Features of letters of recommendation examined by computer software

Thematic category	Words examined
Standout words	Exceptional, best, outstanding, superb, stellar, excellent, phenomenal
Ability	Intelligent, bright, talent, brilliant, competent, smart, gifted
Grindstone	Organized, hardworking, conscientious, diligent
Compassion	Caring, kind, empathy, compassionate

Pearson’s chi-square test was performed to assess possible differences in the presence of each key word by gender, URM, AOA status, USMLE Step 1 score, and match status. For this analysis, the median Step 1 score (240) was used to separate applicants into two groups: “low” scores and “high” scores. For this analysis, a Bonferroni correction was applied to account for multiple comparisons (corrected $\alpha = 0.05/22$, or 0.002). We also performed linear regressions to predict the number of times each thematic category appeared in LORs from the applicant’s gender, URM status, AOA status, USMLE Step 1 score, and match status, again applying a Bonferroni correction (corrected $\alpha = 0.05/20$, or 0.0025).

Results

Of 2,524 applicants, 1 did not report a USMLE Step 1 score and was removed from the study. A total of 2,523 applicants were included in the final analyses. Females comprised 42.2% ($n = 1,064$) of all applicants. The mean USMLE Step 1 score was 238.3 ± 15.8 (median: 240), and the mean USMLE Step 2 score was 243.6 ± 14.4 (median: 246). Among the sample, 479 (19%) were elected into AOA and 1,092 (43.3%) reported either no AOA chapter at their school or unknown AOA status at the time of submission. Of the 387 applicants in the 2018 match cycle who could provide their URM status, 40 (10.3%) self-classified themselves as URMs (**Table 2**).

There were significant differences in USMLE Step 1 and Step 2 scores by URM status, gender, AOA status, and match status. URMs scored significantly lower than non-URMs ($p = 0.001$ for Step 1, $p = 0.015$ for Step 2), females scored significantly lower than males ($p < 0.001$ for Step 1, $p = 0.037$ for Step 2), non-AOA members scored significantly lower than AOA

Table 2 Characteristics of applicants to the ophthalmology residency program at the UCI GHEI

Baseline characteristic	Candidates
<i>N</i>	2,523 ^a
Match 2011, <i>n</i>	281
Match 2012, <i>n</i>	276
Match 2013, <i>n</i>	232
Match 2014, <i>n</i>	302
Match 2015, <i>n</i>	331
Match 2016, <i>n</i>	384
Match 2017, <i>n</i>	352
Match 2018, <i>n</i>	365
Gender (female), <i>n</i> (% respondents)	1,064 (42.2%)
Underrepresented minority, <i>n</i> (% respondents)	40/387
USMLE Step 1 score, mean ± SD (median)	238.3 ± 15.75 (240)
USMLE Step 2, mean ± SD (median)	243.6 ± 16.4 (246)
Applicants with score included in the application, <i>n</i> (%)	1,788 (70.9%)
Alpha Omega Alpha Honor Society	
Yes, <i>n</i> (%)	479 (19%)
No, <i>n</i> (%)	952 (37.7%)
Undetermined at submission or school does not have chapter, <i>n</i> (%)	1,092 (43.3%)

Abbreviations: SD, standard deviation; UCI GHEI, University of California, Irvine Gavin Herbert Eye Institute; USMLE, United States Medical Licensing Exam.

^aApplications of those who failed to match and reapplied in subsequent years were included in analyses as independent applications due to presumed differences in letters of recommendation content by year of submission.

members ($p < 0.001$ for Steps 1 and 2), and unmatched applicants scored significantly lower than matched applicants ($p < 0.001$ for Steps 1 and 2) (►Fig. 2).

The distribution of each of the 22 key words and 4 thematic categories of words are shown in ►Fig. 3; all distributions were heavily positively skewed. Chi-square testing showed differences in the presence of key words in LORs by gender, URM status, USMLE Step 1 score, and match outcome. Females were more likely to be described as “empathetic” ($p = 0.002$), URM status was more likely to be described as “caring” ($p = 0.002$), high Step 1 scorers (≥ 240) were more likely to be described as “outstanding” ($p = 0.002$), and matched students were more likely to be described as “exceptional” ($p = 0.001$), “outstanding” ($p < 0.001$), and “superb” ($p = 0.001$). “Competent” appeared more commonly in low Step 1 scorers ($p < 0.001$) and unmatched applicants ($p = 0.001$) (►Table 3).

The results of the linear regression revealed only the applicant’s AOA status, Step 1 score, and match status were predictive of the number of times key words appeared in LORs (►Table 4). Specifically, standout words appeared

significantly more often in the LORs of AOA members than nonmembers, matched than unmatched applicants, and “high” versus “low” Step 1 scorers ($p < 0.001$ for all comparisons). There was a trend toward female applicants being described with terms from the compassionate category ($p = 0.009$), but this effect failed to cross significance according to our preset α criteria corrected for multiple comparisons ($p < 0.0025$). URM status did not predict the frequency at which each category of key words appeared in LORs.

Discussion

Ophthalmology is among the most competitive specialties in the United States, with most applicants achieving high board scores, several research publications, and excellent clinical grades.¹ Comparing applicants is a difficult task due to few objective metrics beyond USMLE board scores. Additionally, several qualifications in the applications are linked, such as AOA status and high USMLE score. This study of a large set of residency applications suggests that there may exist other interdependencies between application components. In particular, halo effects and other biases may extend into LORs. Moreover, race and gender may affect LOR content.

Our analysis of the baseline characteristics of this sample found significant differences in USMLE Step 1 scores by URM status, gender, AOA status, and match outcome. A high USMLE score is frequently a prerequisite for induction into AOA at most medical schools¹⁶; therefore, the finding that members scored significantly higher than nonmembers is not surprising. Significant difference in USMLE Step 1 scores by match outcome was evident in this cohort, as previously described.⁴ Though a small difference, women scored significantly lower on USMLE Step 1 than men, a finding that has been previously documented.¹⁷ This difference in scores may be attributed to differences in educational backgrounds. More men pursue undergraduate degrees in basic sciences than women, and the USMLE is geared toward the basic sciences.^{17,18} Interestingly, despite this difference in mean USMLE Step 1 scores, there were significantly more women inducted into AOA than men in our sample. URM status scored significantly lower than non-URMs on the USMLE, a finding consistent with previous studies.^{19,20} Disparities in financial and other resources likely play major contributory roles; some studies suggest baseline cognitive differences, but the reasons remain not completely clear.²¹

We identified differences in key words used to describe URM status and females. URM status was more commonly described as “caring,” and females were more commonly described as “empathetic.” However, there was no difference in the frequency at which the categories of key words (standout words, ability, grindstone, or compassion) appeared by URM status or gender. Overall, these data corroborate previous studies documenting differences in LOR and MSPE content by race and gender.^{10–12} The relevance may be negligible, as “empathetic” and “caring” were not more or less likely to appear in the LORs of matched applicants.

An interesting finding was that the key word “competent” was seen more commonly in LORs of unmatched applicants

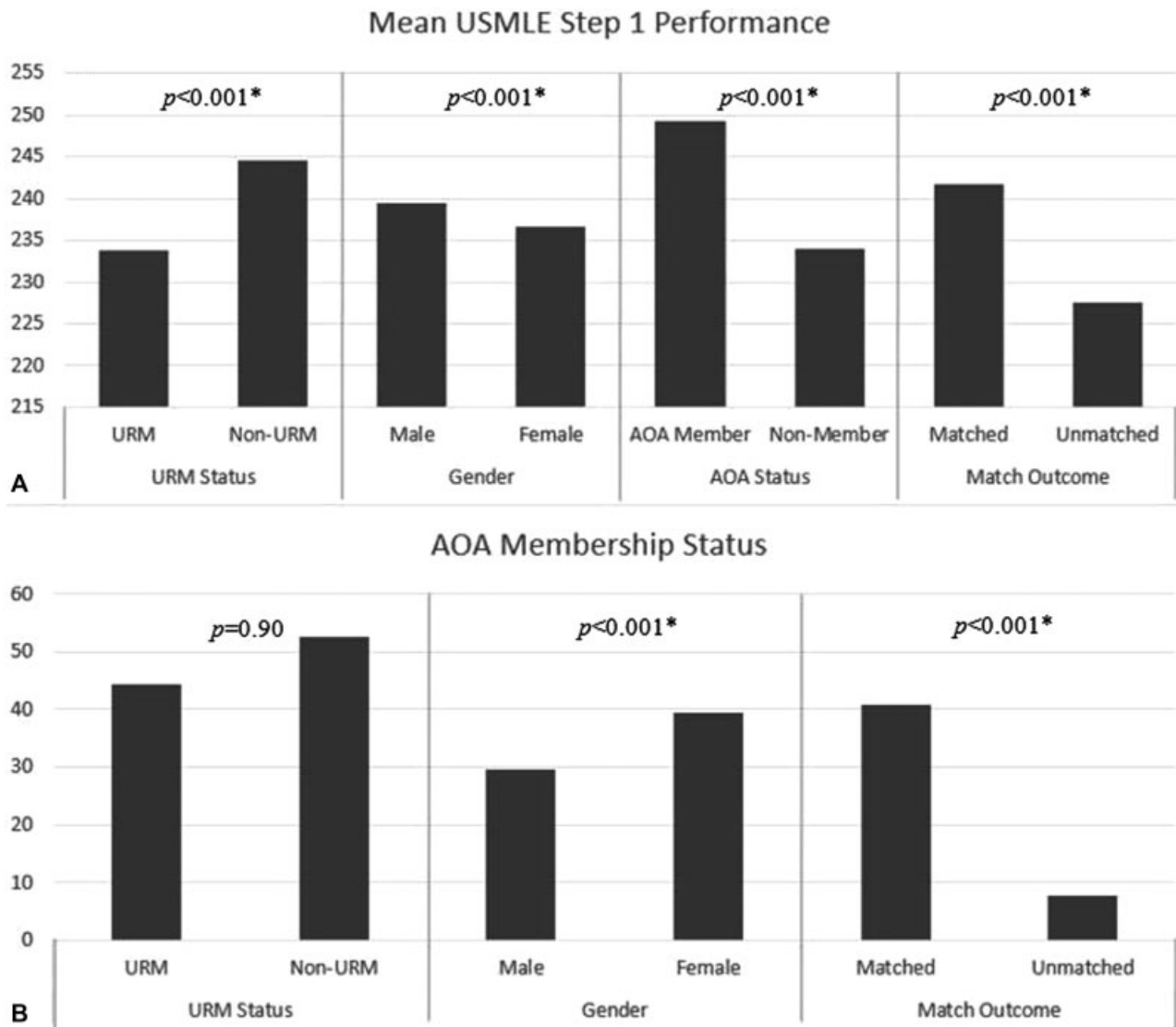


Fig. 2 (A) Mean United States Medical Licensing Exam (USMLE) Step 1 score by baseline characteristics. (B) Proportion of applicants who are Alpha Omega Alpha (AOA) by baseline characteristic. *p*-Values with asterisks represent statistically significant differences ($p < 0.05$). URM, underrepresented minority.

and in those with low Step 1 scores, suggesting that it may not be a good indicator of the strength of an applicant. Ross et al report that Black applicants were more likely to be described as “competent.”¹¹ Our findings may corroborate their suspicions about racial discrimination by evaluators.

Conversely, LORs of high Step 1 scorers were more likely to contain the word “outstanding,” and matched students were more likely to be described as “exceptional,” “outstanding,” and “superb.” Standout words also appeared at a higher frequency in the LORs of matched versus unmatched applicants and AOA members versus nonmembers. Standout word frequency also correlated directly with USMLE Step 1 score. We deduce that there may be a halo effect in which a high Step 1 score or AOA induction may earn a candidate positive key words beyond what these achievements define. We also acknowledge the possibility of the alternative theory that those applicants who are top academic performers are also strong in other dimensions. Further research is warranted.

LORs offer information regarding an applicant that examination scores and clinical grades cannot reveal, including personality traits, work ethic, and interactions with patients. Despite their subjective content, strong LORs have been shown to correlate well with both faculty evaluations and examination scores during residency.^{22,23} At face value, nearly all LORs to residency programs are positive in nature, owing to applicant selection of individuals who they believe will write strong LORs.^{24,25} Furthermore, there is no standardization of LOR length or content, creating difficulties in differentiating the strength of LORs,²⁶ which is the likely basis for the AUPO’s pilot standardized LOR initiative. This study showed that standout words appeared significantly more often in matched applicants compared with unmatched applicants. Key word frequency may offer some objectivity for appraising LORs. It may also offer the ability to stratify LORs from the same letter writer.

This study has several limitations. While the self-reported application data are meant to be truthful, we cannot confirm

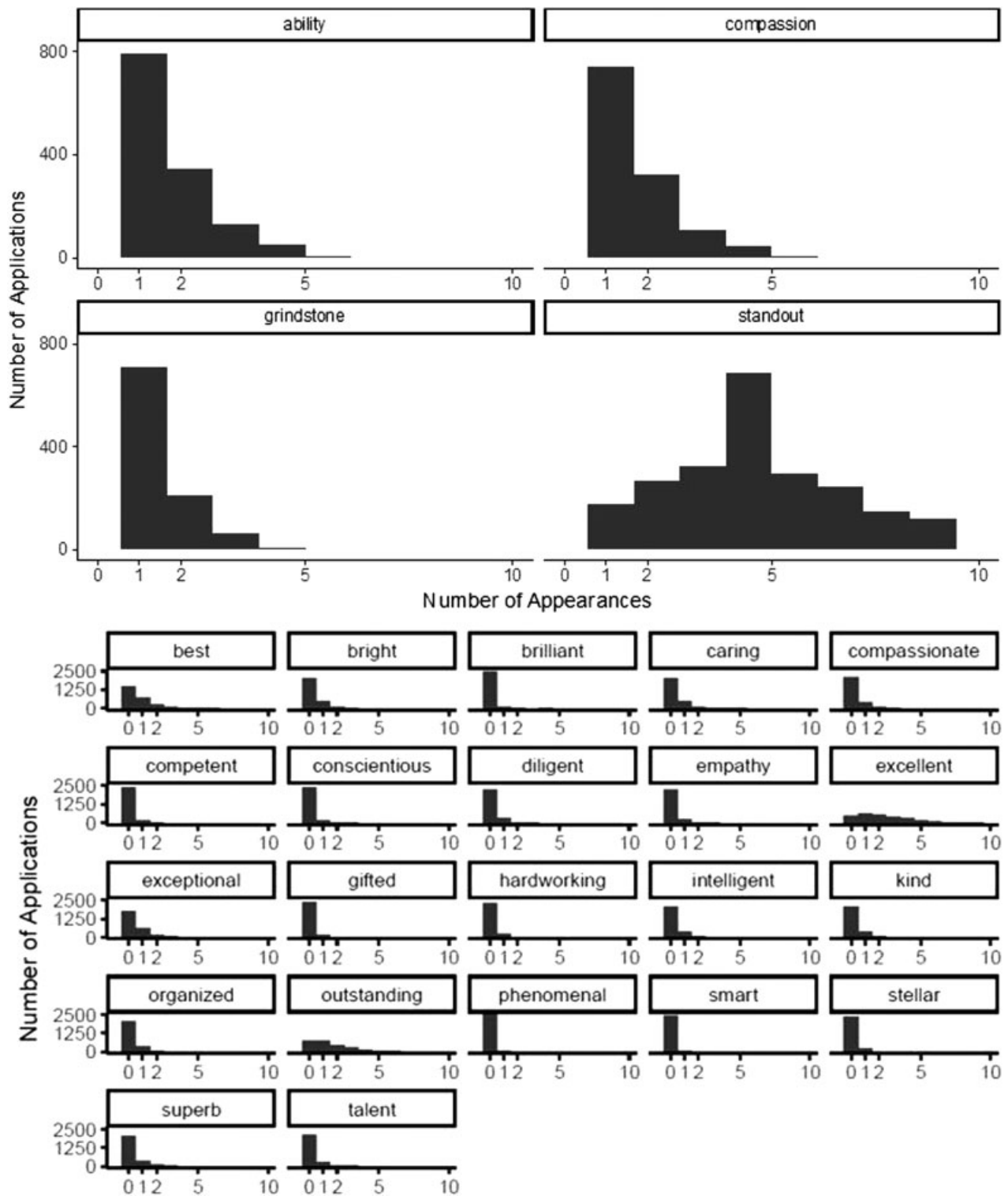


Fig. 3 Frequency histograms of each word and each category.

the accuracy of the information (such as URM status) each applicant provided. Almost half (1,092) of all applicants in our study did not report their AOA status in their application, meaning that they either did not have an AOA chapter at their school or their AOA status was undetermined at the time of submission. Our study was unable to distinguish which of the two reasons AOA status was not reported for each applicant.

However, 96% of all U.S. medical schools have an AOA chapter; therefore, it is likely that most of these candidates simply did not know their AOA status when they submitted their applications. Our study revealed that AOA status may influence descriptive words found in LORs; hence, late induction may disadvantage some members. We acknowledge that many of the 1,092 candidates were able to update

Table 3 Chi-square results of the presence of descriptive words by baseline characteristic

Word categories	Gender		URM		USME Step 1		AOA member		Match outcome		p-Value
	Male (%)	Female (%)	Yes (%)	No (%)	High (%)	Low (%)	Yes (%)	No (%)	Matched (%)	Unmatched (%)	
Standout words											
Exceptional	29	32	48	38	32	29	34	28	32	25	0.001
Best	42	45	48	50	46	41	47	40	45	38	0.004
Outstanding	69	70	75	73	73	67	72	68	72	62	<0.001
Superb	19	21	25	21	20	19	22	19	21	15	0.001
Excellent	81	83	88	84	81	83	82	85	82	82	0.728
Phenomenal	2	2	3	4	2	1	2	1	2	1	0.135
Ability											
Intelligent	18	18	18	20	18	18	20	16	18	17	0.591
Bright	22	22	13	27	24	20	20	23	22	20	0.157
Talented	13	17	20	20	16	15	14	14	15	15	1
Brilliant	3	4	3	4	4	3	2	3	4	2	0.149
Competent	6	6	3	6	4	8	3	7	5	9	0.001
Smart	4	4	8	4	4	4	6	4	4	3	0.421
Gifted	5	6	8	5	4	7	5	5	6	4	0.202
Grindstone											
Organized	17	22	10	16	20	18	20	18	20	16	0.064
Hardworking	9	8	13	13	8	9	9	8	9	8	0.925
Conscientious	8	6	13	6	7	8	7	9	6	9	0.032
Diligent	14	10	10	14	13	12	15	11	13	12	0.566
Compassion											
Caring	19	22	48	24	19	21	18	19	20	21	0.637
Kind	18	16	23	25	17	18	14	15	17	17	0.878
Empathy	9	13	10	14	10	12	13	11	11	11	0.785
Compassionate	18	20	20	24	20	17	19	18	19	17	0.232

Abbreviations: AOA, Alpha Omega Alpha; URM, underrepresented minority; USMLE, United States Medical Licensing Exam.
 Note: proportions are rounded to the nearest percent. Bold-faced p-values represent statistically significant results (corrected α level of significance is 0.002).

Table 4 Linear regression results

Independent variable	Standout		Ability		Grindstone		Compassion	
	Regression coefficient	p-Value	Regression coefficient	p-Value	Regression coefficient	p-Value	Regression coefficient	p-Value
Step 1 score	0.02	<0.001	≤0.001	0.753	0.001	0.256	≤0.001	0.852
Gender ^a	-0.30	0.012	-0.11	0.012	0.01	0.851	-0.11	0.009
URM status ^b	0.07	0.909	-0.17	0.391	-0.08	0.584	0.05	0.805
AOA ^c	0.59	<0.001	-0.02	0.670	0.05	0.217	0.02	0.648
Match status ^d	0.95	<0.001	0.03	0.519	0.01	0.728	0.02	0.624

Abbreviations: AOA, Alpha Omega Alpha; URM, underrepresented minority.

Note: Bold-faced p-values represent statistically significant results (corrected α level of significance is 0.0025).

^aFor gender, 0 represents female and 1 represents male, such that a negative regression coefficient means that the category of key words appeared more frequently in females, and a positive regression coefficient means that they appeared more frequently in males. ^bFor URM status, 0 represents non-URM and 1 represents URM, such that a negative regression coefficient means that the category of key words appeared more frequently in non-URMs, and a positive regression coefficient means that they appeared more frequently in URMs. ^cFor AOA membership, 0 represents nonmember and 1 represents member, such that a negative regression coefficient means that the category of key words appeared more frequently in nonmembers, and a positive regression coefficient means that they appeared more frequently in members. ^dFor match status, 0 represents unmatched and 1 represents matched, such that a negative regression coefficient means that the category of key words appeared more frequently in unmatched applicants, and a positive regression coefficient means that they appeared more frequently in matched applicants.

their AOA status after submitting their applications as they were inducted. However, owing to inconsistencies in timing and method by which AOA was reported, 1,092 candidates were excluded from analyses of AOA.

Our study also reflects only the applicants to the ophthalmology program at the UCI Gavin Herbert Eye Institute and not the national cohort of annual applicants, making it difficult to extrapolate our results to the entire applicant pool. Also, our data abstraction methods using programming script may miss other positive descriptive words and do not capture subtle intonations and other semantics that a reader may glean from reading an LOR in its entirety. Finally, URM status was only reported in the 2018 application cycle, and our LOR analysis was therefore limited to 387 applicants. As future applicants provide their URM status, revisiting these data from a larger sample would be valuable.

In summary, LORs remain one of the most difficult aspects of the application to assess in an objective, meaningful fashion. Biases and halo effects may further confound LORs. An algorithmic approach looking for the presence and frequency of key words may offer insights into the relevance of LORs as well as tendencies of letter writers.

Funding

This study was supported in part by an unrestricted departmental grant from the Research to Prevent Blindness and the Heed Foundation to one of the authors (S. G.).

Conflict of Interest

None.

Acknowledgments

The authors would like to thank Dr. Angele Nalbandian for her assistance in reviewing the literature in the preparation for this study.

References

- SF Match. Ophthalmology Residency Match Summary Report 2018. Available at: https://www.sfmatch.org/PDFFilesDisplay/Ophthalmology_Residency_Stats_2018.pdf. Accessed June 15, 2018
- Nallasamy S, Uhler T, Nallasamy N, Tapino PJ, Volpe NJ. Ophthalmology resident selection: current trends in selection criteria and improving the process. *Ophthalmology* 2010;117(05):1041–1047
- Grubbs JR Jr, Mian SI. Advising students interested in ophthalmology: a summary of the evidence. *Ophthalmology* 2016;123(07):1406–1410
- Loh AR, Joseph D, Keenan JD, Lietman TM, Naseri A. Predictors of matching in an ophthalmology residency program. *Ophthalmology* 2013;120(04):865–870
- SF Match. Ophthalmology Residency Standardized Letter of Reference. Available at: https://www.sfmatch.org/PDFFilesDisplay/OPHTHR_StandardizedLOR.pdf. Accessed June 16, 2018
- Lee AG, Golnik KC, Oetting TA, et al. Re-engineering the resident applicant selection process in ophthalmology: a literature review and recommendations for improvement. *Surv Ophthalmol* 2008;53(02):164–176
- Pololi LH, Civian JT, Brennan RT, Dottolo AL, Krupat E. Experiencing the culture of academic medicine: gender matters, a national study. *J Gen Intern Med* 2013;28(02):201–207
- Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agent and communal differences. *J Appl Psychol* 2009;94(06):1591–1599
- Schmader T, Whitehead J, Wysocki VH. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles* 2007;57(7-8):509–514
- Trix F, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse Soc* 2003;14(02):191–220
- Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. *PLoS ONE* 2017;12(08):e0181659
- Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A linguistic analysis of Medical Student Performance Evaluations. *Acad Med* 2011;86(01):59–66
- Lachman SJ, Bass AR. A direct study of halo effect. *J Psychol* 1985;119(06):535–540

- 14 Gibson JL, Gore JS. Is he a hero or a weirdo? How norm violations influence the halo effect. *Gend Issues* 2016;33(04):299–310
- 15 Messner AH, Shimahara E. Letters of recommendation to an otolaryngology/head and neck surgery residency program: their function and the role of gender. *Laryngoscope* 2008;118(08):1335–1344
- 16 Gauer JL, Jackson JB. Association between the Medical College Admission Test scores and Alpha Omega Alpha Medical Honors Society membership. *Adv Med Educ Pract* 2017;8:627–632
- 17 Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. *Acad Med* 2008;83(10, Suppl):S58–S62
- 18 Case SM, Becker DF, Swanson DB. Performances of men and women on NBME Part I and Part II: the more things change.... *Acad Med* 1993;68(10, Suppl):S25–S27
- 19 Lypson ML, Ross PT, Hamstra SJ, Haftel HM, Gruppen LD, Colletti LM. Evidence for increasing diversity in graduate medical education: the competence of underrepresented minority residents measured by an intern objective structured clinical examination. *J Grad Med Educ* 2010;2(03):354–359
- 20 Edmond MB, Deschenes JL, Eckler M, Wenzel RP. Racial bias in using USMLE step 1 scores to grant internal medicine residency interviews. *Acad Med* 2001;76(12):1253–1256
- 21 Tekian A, Han Y, Hruska L, Krainik AJ. Do underrepresented minority medical students differ from non-minority students in problem-solving ability? *Teach Learn Med* 2001;13(02):86–91
- 22 Xierali IM, Nivet MA, Wilson MR. Current and future status of diversity in ophthalmologist workforce. *JAMA Ophthalmol* 2016;134(09):1016–1023
- 23 Brothers TE, Wetherholt S. Importance of the faculty interview during the resident application process. *J Surg Educ* 2007;64(06):378–385
- 24 Stohl HE, Hueppchen NA, Bienstock JL. The utility of letters of recommendation in predicting resident success: can the ACGME competencies help? *J Grad Med Educ* 2011;3(03):387–390
- 25 Daly KA, Levine SC, Adams GL. Predictors for resident success in otolaryngology. *J Am Coll Surg* 2006;202(04):649–654
- 26 Rim Y. How reliable are letters of recommendation? *J Higher Educ* 1976;47(04):437–445